2M-09

RoboCupSoccerにおける実数値GAを用いたサッカーAIの学習

水島　諒†　　穴田　一†

東京都市大学†

1. 研究背景及び目的

近年，「ゲーム AI」の開発が盛んに行われている．例えば，チェスや将棋，囲碁といったゲームが挙げられる．そして，これらのゲームにおいてはAI が人間のチャンピオンに勝利するといった事も起きている．また，RoboCup と呼ばれる，自律型ロボットによるサッカーの世界大会が毎年行われている[1]．RoboCup とは，西暦 2050 年迄にサッカーの世界チャンピオンチームに勝てる，自律型ロボットのチームを作ることを目標とした大会である．このRoboCup には 5 つのリーグがあり，リーグごとに異なる特徴がある．本研究では 5 つのリーグのうち，各選手がそれぞれ思考し，人間のような戦術的なサッカーが行われている 2D リーグを扱う．2D リーグにおけるこれまでの多くの研究では，戦況に応じた意思決定ができているとは言い難い．そこで，本研究では実数値 GA を用いて戦況に応じた選手のAI を学習し，その有効性を確認した．

2. 既存研究

秋山は RoboCup の 2D リーグ（高さの概念がない）で使用可能な agent2d(Ver 3.1.1) というチームモデルを公開している[2][3]．2D リーグで用いるフィールドは，中央を原点とし，長辺方向を x 軸，短辺方向を y 軸とした直交座標系で，x 軸は自分のゴール側が負の値となっている．フィールドの大きさは長さ $105m$，幅 $68m$ である．

agent2d は，ボールを保持しているエージェントの意思決定を，ゲーム木探索を参考に提案した「Chain Action モデル」を用いて行う．そして，ボールを保持していないエージェントの意思決定は，提案したボールの位置から移動場所を決定するという「フォーメーションシステム」を用いて行う．

3. 提案モデル

本研究では，ボールを保持している選手だけでなく，ボールを保持していない選手に関しても ChainAction モデルを適用することにより，戦況に応じた最適な意思決定を実現する．ChainAction モデルは，実行可能な行動についてそれぞれ戦況に応じた評価を行い，意思決定を行う．

しかし，これまでの研究では意思決定を行うための評価値 V を経験則で構築してきた．そこで本研究では，実数値 GA を用いて重みや閾値を学習させる．

3.1 ボール保持者の意思決定方法

3.1.1 ChainAction モデルの概要

現在の全選手の配置を初期状態とし，ボール保持者が考えたパスとドリブルの行動によって，2 手先まで考え，どのような展開パターンがあるかをツリー構造で表す．そして，その全ての展開を評価し，最も評価が高い展開を選択する．

3.1.2 ChainAction モデルに用いる評価値

展開を評価するために用いる評価値はパスとドリブルで異なると考えたため，行動 a（パス，ドリブル）の評価値 V^a を次のように定義した．

$$V^a = \omega_0 U_0 + \sum_{i=1}^{4} \omega_i^a U_i \tag{1}$$

ここで，ω_0 は全行動共通の評価項 U_0 の重み，ω_i^a は行動 a の評価項 U_i の重みを表す．

全行動共通の評価項 U_0 には次のようなものを定義した．

$$U_0 = \frac{34.0 - |y_b|}{34.0} \qquad \text{if } (x_{th} < x_b) \tag{2}$$

ここで，y_b は行動後のボールの y 座標，x_{th} は x 軸の閾値を表す．2D リーグでは高さの概念が無く，敵陣内のサイドに進んだとしても頭上を通すパスができない．そのため，両サイドを進んだとしても，行動の選択肢が狭まるだけであると考えたため，敵陣内のサイドより中央の方が高評価になるように調整した．

行動 a の評価項 U_i には次のように定義した．

$$U_1 = \frac{x_b + 52.5}{105.0} \tag{3}$$

$$U_2 = \frac{\max\{0.0,\ dist_{th1}^a - dist_{b,g}\}}{dist_{th1}^a} \tag{4}$$

$$U_3 = \frac{dist_{l,o}}{20.0} \tag{5}$$

$$U_4 = \frac{\max\{0.0, dist_{s,b} - dist_{th2}^a\}}{dist_{s,b}} \tag{6}$$

ここで，x_b は行動後のボールの x 座標，$dist_{th1}^a$ は行動後のボールとゴールとの距離である $dist_{b,g}$ の閾値，$dist_{l,o}$ はボールの移動軌跡から一番近い敵との距離，$dist_{th2}^a$ は今の位置と行動後のボールとの距離である $dist_{s,b}$ の閾値を表す．そして，全ての評価項がおおよそ 0 以上 1 以下の範囲をとるように調整した．

Soccer AI using Real-coded GA for RoboCupSoccer
† Ryo Mizushima, Hajime Anada, Tokyo City University

Copyright ©2018 Information Processing Society of Japan.
All Rights Reserved.

3.2 ボール非保持者の意思決定方法

3.2.1 ChainAction モデルの適用

現在の全選手の配置を初期状態とし，ボール非保持者が考えた 20 方向（各 5 パターン）と今の位置の計 101 パターンの移動によって，局面がどのように変化するかをツリー構造で表す．そして，その全ての展開を評価し，最も評価が高い展開を選択する．

3.2.2 ChainAction モデルに用いる評価値

移動のための評価値 V は次のように定義した．

$$V = \sum_{i=1}^{7} \omega_i U_i \tag{7}$$

ここで，ω_iは評価項U_iの重みを表す．

評価項 U_i には以下のような 7 つの評価項目を用意した．

$$U_1 = \frac{x_s + 52.5}{105.0} \tag{8}$$

$$U_2 = \frac{\max\{0.0, dist_{th1} - dist_{s,g}\}}{dist_{th1}} \tag{9}$$

$$U_3 = \frac{dist_{l,o}}{20.0} \tag{10}$$

$$U_4 = \frac{ang_{s,g}}{180.0} \tag{11}$$

$$U_5 = \min\left\{1.0, 1.0 - \frac{dist_{s,h} - Min_{th1}}{52.5 - Min_{th1}}\right\} \tag{12}$$

$$U_6 = \frac{\min\{dist_{th2}, dist_{s,t}\}}{dist_{th2}} \tag{13}$$

$$U_7 = \min\left\{1.0, 1.0 - \frac{dist_{s,b} - Min_{th2}}{52.5 - Min_{th2}}\right\} \tag{14}$$

ここで，x_s は移動候補の x 座標，$dist_{s,g}$ は移動候補とゴールとの距離，$dist_{l,o}$ は移動候補とボールとのライン上から一番近い敵との距離，$ang_{s,g}$ は移動候補とゴールとのなす角，$dist_{s,h}$ は移動候補と既存研究の移動場所の決定方法である「フォーメーションシステム」により決定した移動先との距離，$dist_{s,t}$ は移動候補から一番近い敵との距離を表す．$dist_{th1}, dist_{th2}, Min_{th1}, Min_{th2}$ はそれぞれ閾値である．また，評価項目 U_i はおおよそ 0 以上 1 以下の範囲をとるように設計した．

3.3 評価値を計算するために必要なパラメータ

(1)式より重みは 9 個，(2)~(6)式より閾値は 5 個，(7)式より重みは 7 個，(8)~(14)式より 4 個となり，計 25 個である．そして，選手の意思決定はポジションごとに異なると考え，6 つのポジションごとにパラメータを用意し，総パラメータ数は 150 個とした．

3.4 実数値 GA の適用方法

本研究では用意した 4 チームに勝てるように学習を行った．チームの評価は用意した 4 チームとそれぞれ 50 回ずつ対戦させ，次式で表される適合度 G を用いて行う．

$$G = \frac{\sum_{i=1}^{4}\sum_{j=1}^{20}(P_{ij} + D_{ij}/1000)}{4 \times 20} \tag{15}$$

ここで，P_{ij} はチーム i の j 回目の勝ち点であり，勝利なら 3，引分なら 1，敗北なら 0 を表す．D_{ij} はその試合の得失点差を表す．そのため，チームの評価は 80 試合の勝ち点の平均を用い，勝ち点の平均が同じ際は得失点差を考慮する．

実数値 GA の流れは以下の通りである．

I. 初期世代
150 個のパラメータをランダムに決定したチームを 12 チーム作成する．用意した 4 チームとそれぞれ 50 試合対戦し，(15)式を用いて適合度 G を求める．

II. 交叉
ランキング選択[4]によって 2 チームを選択することを繰り返し，12 組作成する．そして，1 組ごとに 2 チームで BLX-α[5]を用いて交叉を行い，新規チームを作成し，計 12 チームを作成する．

III. 突然変異
新たに作成したチームのパラメータ全てに 3%の確率で，指定されたパラメータ範囲内のランダムな値に変更する．

IV. 適合度計算
残存したチームと新たに作成したチームの計 24 チームを，それぞれ用意した 4 チームと 50 試合ずつ対戦し，(15)式を用いて適合度 G を求める．

V. 選択
エリート戦略[4]に基づき現在の世代における上位半分の 12 チームを次の世代に残す．

VI. 終了条件
II～V を指定した世代まで繰り返す．

4. モデルの評価

発表時に学習結果と考察を述べる．

参考文献

[1] “ロボカップ日本委員会”, http://www.robocup.or.jp/original/about.html

[2] 秋山 英久，“アクション連鎖探索によるオンライン戦術プランニング”，人工知能学会研究会資料，SIG-Challenge-B101-6， pp.23-28，2011.

[3] 秋山 英久，野田 五十樹，“RoboCup サッカーシミュレーションにおけるエージェント配置手法の提案“，情報処理学会研究報告ゲーム情報学，Vol.2007，No.20，pp.9-16，2007

[4] 佐藤 浩，小野 功，小林 重信，“遺伝的アルゴリズムにおける世代交代モデルの提案と評価”，人工知能学会誌，Vol.12， No.5，pp.734-743，1997.

[5] Eshelman, L.J，“Real-Coded Genetic Algorithm and Interval Schema ta”，Foundations of Genetic Algorithm 2，pp.187-202，1993.

Copyright ©2018 Information Processing Society of Japan.
All Rights Reserved.